Culture-specific communication management for virtual agents

Endrass, Birgit; Rehm, Matthias; André, Elisabeth

Published in:
International Conference on Autonomous Agents : Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Culture-specific Communication Management for Virtual Agents

Birgit Endrass
Augsburg University
Eichleitnerstr. 30
D-86135 Augsburg, Germany
+49 (821) 598 23 38
endrass@informatik.uni-augsburg.de

Matthias Rehm
Augsburg University
Eichleitnerstr. 30
D-86135 Augsburg, Germany
+49 (821) 598 23 43
rehm@informatik.uni-augsburg.de

Elisabeth André
Augsburg University
Eichleitnerstr. 30
D-86135 Augsburg, Germany
+49 (821) 598 23 41
andre@informatik.uni-augsburg.de

ABSTRACT
Human interaction depends on several individual factors such as personality, social relations, age or gender. But also the society we live in influences our behaviour. Thus culture affects the way communication is led. As virtual agents interact in a more and more human-like manner, culture-specific behaviour should also be taken into account. In this paper, we investigate communication management as one aspect of communication. Our findings in culture related differences are based on a video corpus that was recorded in Germany and Japan as well as on findings described in the literature. To this end, the use of pauses in speech as well as the occurrence of overlapping speech was analyzed and integrated into a demonstrator using virtual agents. In a preliminary study, we investigated whether subjects perceive a difference between agent dialogs that are in line with culture-specific findings and agent dialogs that are not.

Categories and Subject Descriptors
I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence - *Intelligent agents*; I.6.7 [**Simulation and Modeling**]: Model Development

General Terms
Human Factors, Theory.

Keywords
Virtual agents, behavior simulation, culture, communication management.

1. INTRODUCTION
When people communicate they do not need to think about the management of their conversation. Tasks such as turn taking or pauses in speech are solved automatically without being aware of it. So, for example one communication partner starts talking when the other one stops. But these mechanisms pose a great challenge to computer based dialog systems.

Cite as: Culture-specific Communication Management for Virtual Agents, Birgit Endrass, Matthias Rehm, Elisabeth André, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009),* Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. 281–288

Communication depends on individual factors such as personality, age, gender or personality on the one hand and on collective factors such as culture on the other hand. Models of personality or the development of personal relations have already been successfully integrated in systems using embodied conversational agents (ECAs). Although ECAs communicate in a more and more human-like and personalized manner, so far little effort has been made to integrate common context such as cultural background into their behaviour. In this paper we focus on culture related differences in communication management as one aspect of culture specific behaviour. Thus pauses in speech as well as the usage of overlapping speech have been observed in different cultures (Germany and Japan).

Using pauses in speech in dialog simulation systems is not a recent topic at all. Sometimes silence in speech arises due to a lack of celerity in the speech components and thus appears to be distracting for the user. In [1] Gudny Ragna et al. describe an attempt in using machine learning in order to build an agent that learns proper turn taking during interaction and adjusts its behaviour to its communication partner. Therefore optimal pause durations are learned to minimize speech overlaps. But as the authors state the shortest duration of silence between turn is not always the most efficient technique. Again the comfortableness with pauses in speech or overlapping speech is dependent from several factors such as personality or culture.

Another approach to learning turn taking has been presented by Sato et al. [2]. They analyzed a large corpus of human-human conversation in order to train a decision tree. Their results show that semantic and syntactic categories are also important features to handle turn taking besides the usage of silence in speech.

In other systems pauses in speech are explicitly used to handle tasks such as turn taking. Sidner and colleagues [3] for example, developed a model of engagement for a conversational robot, based on an analysis of human-human conversation. Engagement "is the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake". The appropriate use and correct interpretation of engagement signals are necessary prerequisites for the success of an interaction. In particular, pauses are used to recognize inattentiveness of the user, which encourages the robot to show engagement behaviour.

Another approach of using pauses in speech is described in [4], where silence is used for grounding behaviour of ECAs. Cassell and colleagues present a Real Estate Agent (REA) that acts in the

function of a virtual realtor. In Smalltalk situations, she gains information about the users` preferences when buying a house. In [5], Cassell states that short pauses in speech lead to feedback behaviour. Thus, the REA agent nods her head or emits a paraverbal (such as "Mmhmm") or a short statement (such as "Okay") as reaction to short pauses in the user's speech.

Based on this work, Nakano and colleagues [6] developed a grounding model for the kiosk agent Mack that provides route descriptions for a paper map. The agent uses verbal and nonverbal grounding acts to update the state of the dialogue. The authors state that pauses influence the choice of subsequent actions.

Traum and Heeman [7] also consider grounding behaviour in dialogues. They examine the co-occurrence between turn-initial grounding acts and utterance unit signals, e.g. prosodic boundary tones and pauses. The authors categorized silence into two groups: short silence (less than half a second) and long silence (longer than half a second). In this vein correlations with boundary tones and relatedness markings were analysed. As a result they state that long pauses are positively related with the previous utterance being grounded and that those pauses seem to be an indicator of utterance unit completion.

Nakanishi and colleagues [8] describe a helper agent that plays the role of a party host in a virtual meeting space where different cultures meet. In this system silence is used to detect conversations that are going badly. When the helper agent locates a pause in speech, it directs a series of yes/no questions to both conversation partners in order to find a topic that is interesting for both. Although the agent is developed to help in intercultural encounters, the length of silence that initiates the agent is not adapted to culture. After analysing their results, the authors state that an adaption to the user's cultural background would make the agent more efficient. But so far a cultural aspect in the use of silence has not been taken into account.

The flow of a conversation depends on several variables such as personality or personal relationships between the interlocutors. Thus a dialog between two friends is completely different than a conversation between strangers. This phenomenon has been analysed by Cassell et al [9] in order to build a computational model of the role of long-term relationships in language use between humans and embodied conversational agents. Therefore they analysed differences in features such as eye-gaze or head nods. Interestingly the authors state that details such as leaning in towards one another, laughing, telling jokes at one another's expense, and interrupting each other differs from culture to culture. Again, the cultural background of communication partners that engage in a conversation plays a curial role in how the conversation is managed.

The work described above shows that there is a need for adapting a characters behavior to the cultural background of the interaction partner. In our work we focus on culture-specific aspects of communication management in order to simulate dialogs with virtual agents that communicate according to a given cultural background. As a starting point we concentrate on differences between Western and Asian cultures.

This paper is organized as follows: First we give some definitions of culture and describe ways to distinguish different cultural groups. According to these classifications we introduce differences in communication management, namely the usage of pauses in speech and overlapping speech. In Section 3 we describe a corpus study, in order to ground tendencies described in the literature in empirical data. Therefore the CUBE-G corpus was analysed in respect of culture related differences in communication management. This video corpus war recorded in the two cultures Germany and Japan. In Section 4 we describe the integration of our findings into a demonstrator using virtual agents. Afterwards we present an evaluation study where we tested whether subjects perceive a difference between culture-specific dialogs that are in line with observations made for their own cultural background and agent dialogs that are not. In Section 5 we discuss our results and describe our future work in this research field.

2. CULTURE AND COMMUNICATION MANAGEMENT

Before integrating culture as a parameter into a computer-based system, a clear definition needs to be found, which is not easy to solve as there are many fuzzy notations about culture. In order to build up a computational model of culture it is necessary to categorize cultures and to measure the impact of belonging to a certain cultural group on behaviour.

There are different approaches to distinguish cultures. The most precise definitions (and thus the most appropriate ones for our purposes) describe culture as a dimensional model or group cultures according to interaction styles.

Hofstede [10] for example explains culture as a dimensional concept. His theory is based on a broad empirical survey in which over 20 different cultures were categorized into a five dimensional model. Each dimension contains two extreme sides, for which he clearly defines stereotypical behaviour norms. He defines a given culture as a point in a five-dimensional space, according to the dimensions. These dimensions are:

1. *Hierarchy*: This dimension describes the extent to which different distribution of power is accepted by the less powerful members.
2. *Identity*: Here, the degree to which individuals are integrated into a group is defined. On the individualist side ties between individuals are loose, and everybody is expected to take care for himself. On the collectivist side, people are integrated into strong, cohesive ingroups.
3. *Gender*: The gender dimension describes the distribution of roles between the genders. In feminine culture the roles differ less than in masculine cultures, where competition is rather accepted and status symbols are of importance.
4. *Uncertainity*: The tolerance for uncertainty and ambiguity is defined in this dimension. It indicates to what extent the members of a culture feel either uncomfortable or comfortable in unstructured situations which are novel, unknown, surprising, or different from usual.
5. *Orientation*: This dimension distinguishes long and short term orientation. Values associated with long term orientation are thrift and perseverance whereas values associated with short term orientation are respect for tradition, fulfilling social obligations, and saving one's face.

According to Hofstede, behaviour varies with positioning on these five dimensions. Regarding communication management the identity dimension is of special interest. The two extreme sides are individualism and collectivism, thus cultures can be divided into a collectivistic and an individualistic group. While most Western cultures belong to the individualistic side, most Asian cultures can be found in the collectivistic group. In Hofstede´s survey over 20 different cultures were rated. Germany lies on the individualistic side of this dimension, whereas Japan is a collectivistic culture. This classification affects the management of conversations, in particular the usage of pauses in speech. In [11] Hofstede states, that in collectivistic cultures silence may occur in conversations without creating tension, which does not hold true for individualistic cultures.

Another distinction of cultural groups is done in [12]. Following Hall, Ting-Toomey distinguishes high- and low context communication cultures. In high context communication little is encoded explicitly and the conversation relies mainly on physical context. Besides verbal utterances, meaning is transported through context (e.g. social roles or positions) or nonverbal clues (e.g. pauses, silence and prosody). Thus, interlocutors are expected to "read between the lines" in order to encode the whole meaning of a verbal message.

In contrast low context communication explicitly codes information. Therefore clear descriptions, unambiguous communication and a high degree of specificity are required. Thus the speaker is expected to build up clear messages that can be understood easily without the need to encode other aspects of communication, such as nonverbal behaviour or pauses in speech.

Again, with this classification a line between Western and Asian cultures can be drawn. While most Western cultures use low context communication, most Asian cultures use high context communication. In [12] Germany is mentioned as one of the most extreme low context cultures. Japan, on the other hand, is named to be on the extreme high context side.

Following Hall (1983), Ting-Toomey [12] claims that silence serves as a critical communication-device in Japanese communication patterns. Pauses reflect the thoughts of the speaker and can contain strong contextual meaning. In contrast, in European conversations pauses are often sensed as unpleasant.

This distinction with its corresponding usage of silent traces in speech is in line with the tendencies described by Hofstede. Thus, in our empirical study (see Section 3), we expected to observe pauses in speech more frequently in the Japanese conversations than in the German ones.

Another analysis describing styles of communication management in different cultural groups has been done in [13]. Cultures are divided into three categories: Anglo Saxon (Western), Latin and Oriental (Asian). The authors describe Western societies as verbal cultures, where members become nervous and uneasy once they stop talking. Furthermore silence might be seen as "failure to communicate". Vice versa in Asian cultures pauses in speech are a sign of respect, where the listener takes some time to process the information given by the speaker.

Overlapping speech is often thought of as interrupting the conversation partner in the sense of *breaking in on someone*. But when two people talk at the same time this does not have to be in an unfriendly manner. Socalled feedback behaviour, e.g. acknowledging what the conversation partner is saying, is often performed while it is still the interlocutor's talking turn. In [14] the semantics and pragmatics of linguistic feedback is investigated. The authors distinguish different types of reactions according to their communicative function:

1. *Contact*: willingness and ability to continue the interaction;
2. *Perception*: willingness and ability to perceive expression and message;
3. *Understanding*: willingness and ability to understand expression and message;
4. *Attitudinal reactions*: willingness and ability to give other attitudinal reactions to expression, message, or interlocutor.

But the usage as well as the perception of overlapping speech is dependent on culture. Interruptions that are meant for taking the speaking floor are often seen as impolite in Western cultures, whereas in Latin cultures these interruptions occur more frequently and are interpreted as interest in what the conversation partner is saying [13]. In Japanese conversations overlapping speech has mainly another function. Communication partners explicitly communicate that they are listening (hai hai) [12], which matches Allwood's reaction type of understanding [14]. As this is less common in Western cultures, we expected to find more overlapping speech in the Japanese conversations than in the German ones in our empirical study (see Section 3). As we stated above, overlaps are mainly used to acknowledge in Japanese conversations. Thus, we assume that overlapping parts of speech in Japanese conversations are short but frequent.

3. EMPIRICAL STUDY

As we stated above, our aim is to build up a multiagent system that simulates culture specific dialogs. For a first demonstrator we want to focus on differences in communication management in Asian versus Western cultures. Findings in literature suggest that in stereotypical Asian conversations more pauses in speech are used than in Western dialogs. Furthermore we assume that Asian dialogs contain more overlapping speech, as signalling understanding is more common in Asian cultures.

In order to gain some deeper insight in how communication is managed in different cultures we analysed the video corpus collected for the CUBE-G project [15].

For the acquisition of this corpus, three prototypical interaction scenarios were videotaped in one Western culture (Germany) and one Asian culture (Japan). Figure 1 shows two examples within the first scenario (first time meeting) in the two participating cultures (upper: Germany; lower: Japan). In this vein around 20 hours of video material were collected, and more than 20 students participated from each culture. To ensure a high control over the recordings, subjects interacted with actors whom they did not know in advance. At the beginning of the experiment, participants were asked to get acquainted with each other as a preparation for the task they had to solve later. During this time recording already started. Due to its prototypical nature, our analysis of communication management started with this scenario.

Figure 1. Example interaction in the empirical study (upper: Germany; lower: Japan)

For the CUBE-G project culture-specific nonverbal behaviour such as posture or expressivity of gestures was analysed (see Figure 1 for stereotypical body postures).

In this work we focus on communication management as a condition between verbal and nonverbal behaviour. As a starting point we annotated eight German and eight Japanese first time meetings using the Anvil tool [16]. To avoid gender effects the behaviour of four male and four female subjects was observed in each culture, with all gender combinations (male x male; female x female; male x female).

Paused in speech can be divided into filled and silence pauses [17]. During a filled pause, sounds like *uhmm* and *ahhm* might occur as well as nonverbal behaviours like head nods or gestures. In comparison a silent pause is, as the name predicts, completely silent.

For our analysis we only took into account those time spans in which neither the subject nor the actor spoke (silent pauses plus filled pauses that were filled by nonverbal clues). To sort out very brief pauses (like those while breathing between sentences) we only observed pauses that last for more than one second. In a later analysis we restricted to pauses that last over 2 seconds as these are silences that last for long enough to create tension in Western cultures (see Section 2). Please note that pauses over 2 seconds are also included in those that last for more than 1 second.

Comparing the usage of pauses in speech between the two cultures, the results are promising. On average we found 7.1 pauses that lasted for more than one second, and 1.3 pauses on average that lasted for more than 2 seconds in the German videos (please note that each video lasted for approximately 5 minutes). In the Japanese videos we observed 31 pauses on average that lasted over 1 second and 8.4 pauses that lasted for more than 2 seconds. Table 1 shows an overview of the average usage of pauses in German and Japanese videos as well as the average usage of pauses per minute. Using the t–test for statistical analysis, we achieved significance for both, pauses that last for more than 1 second ($p < 0.001$) and pauses that last for more than 2 seconds ($p < 0.001$). Figure 2 graphically shows the distribution of short (more than 1 second) and long pauses (more than 2 seconds) in the analysed videos within the two cultures Germany and Japan.

Table 1. Average occurrence of pauses in speech in German and Japanese conversation

Pauses	Germany	Japan
> 1 sec (per video)	7.1	31
> 2 sec (per video)	1.3	8.4
> 1 sec (per minute)	1.4	6.2
> 2 sec (per minute)	0.2	1.7

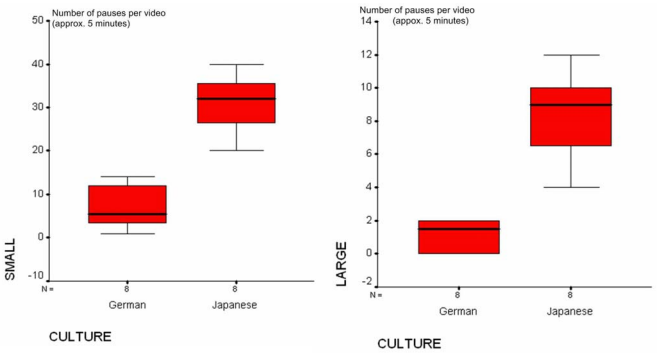

Figure 2. The usage of short (left) and long pauses (right) in speech in the two cultures Germany and Japan.

To ensure that our results were evoked by culture and not by other factors such as gender, we performed an inter cultural analysis as well. Thus we compared female with male subjects and mixed versus same gender constellation within the cultures. For statistical analysis we used the t-test. Comparing female and male subjects in Germany significant results were neither achieved for short pauses ($p = 0.748$) nor for long pauses ($p = 0.750$). Additionally we received no significant results for mixed gender combinations compared to those where both subjects had the same gender (short pauses: $p = 0.795$; long pauses: $p = 0.578$). The same analysis was done for the Japanese videos. On the analogy of the German dialogs we did not achieve significance comparing pauses done by female and male subjects (short pauses: $p = 0.770$; long pauses: $p = 0.252$). In the same way, we did not receive significant results by comparing mixed gender conversations with same gender dialogs in the Japanese videos (short pauses: $p = 0{,}473$; long pauses: $p = 0.425$). In summary our analysis revealed no significant results comparing differences in the usage of pauses in speech between the genders or different gender combinations. In comparison, we achieved significance comparing the two cultures Germany and Japan for both short and long pauses.

Regarding culture related differences in the usage of overlapping speech the results are less obvious. Time spans where both conversation partners spoke at the same time were analyzed. Thus, pragmatics (such as using overlaps for feedback behavior) was not yet taken into account in our preliminary study. For our analysis

we observed the same sixteen videos as described above (8 Japanese videos, 8 German videos, with 4 female and 4 male subjects for each culture within the first time meeting scenario). In contrast to pauses in speech, we took into account all overlaps, more precisely short overlaps as well (less than a second). Later we took a closer look at overlaps in speech that last for more than 1 second and 0.5 seconds respectively, as these are traces where interlocutors do actually speak at the same time. Table 2 shows an overview of the average occurrence of overlapping speech in the German and Japanese videos, as well as the average occurrence per minute. On average we observed 6 overlaps per minute in German conversations, whereas in Japanese dialogs we found 9 overlaps per minute on average. Comparing the frequency of overlapping speech in the two cultures Germany and Japan we achieved significant results for all overlaps ($p = 0.04$). We did not receive significant results for overlaps that last for more that 0.5 seconds ($p = 0.31$) and 1 second ($p = 0.12$) comparing the cultures, although we observed more overlaps in Japanese conversations for both lengths.

Table 2. Average frequency of overlapping speech in German and Japanese conversation

Overlaps	Germany	Japan
> 0 sec (per video)	32.1	46.6
> 0.5 sec (per video)	12.4	14.5
> 1 sec (per video)	2.6	4.3
> 0 sec (per minute)	6.4	9.3
> 0.5 sec (per minute)	2.5	2.9
> 1 sec (per minute)	0.5	0.9

These results suggest that very short overlaps in speech are used in Asian conversation to confirm understanding. This is also in line with findings described in literature that states that acknowledging in the sense of "I am hearing you" is more common in Asian conversations than in Western ones.

4. EVALUATION STUDY

In order to display and evaluate our findings in culture-related communication management differences, we use the Virtual Beergarden, which represents a meeting place where virtual agents interact. Within the CUBE-G project culture-specific virtual agents were created and nonverbal behaviour was investigated [15]. Figure 3 shows two scenarios in the Virtual Beergarden (upper: German agents; lower: Japanese agents).

The characters' appearance (skin, hair or shape of the face) as well as their nonverbal behaviour has been adapted to their cultural and ethnic background. Differences in nonverbal behaviour may manifest themselves in obvious ways, for example by culture-specific gestures (such as a bow for the Japanese greeting) or typical body postures for a given culture (see Figure 3). To show culture-related differences in nonverbal behaviour in a more subtle way, we modeled different ways of performing an action [18]. For instance, the speed, rhythm or spatial extension of a gesture may be varied.

In order to investigate whether subjects perceive a difference between agent dialogs that are in line with culture-specific findings correlated to their own cultural background and agent dialogs that are not, we designed an online evaluation following [18]. To this end we recorded several videos containing a pair of agents that communicate with each other. In their conversation either the usage of pauses or the frequency of overlaps or both is done in a prototypical German or prototypical Japanese way.

Figure 3. Culture specific agents in the Virtual Beergarden application (upper: Germany; lower: Japan)

To avoid side effects aroused by different gender combinations in the agent conversations or depending on the subjects' preferences we decided to show mixed gender combinations in the videos. Thus one female and one male agent interacted with each other.

In another study, we ascertained that the choice of topics as well as the way topics were ordered depended on culture and did affect user perceptions of scenarios with virtual agents [19]. As our focus was on communication management in the preliminary evaluation study described in this paper, we had to assure that subjects are not distracted by the semantics of the dialog or any preferences correlated to the topics discussed in the conversation. Thus, the agents spoke Gibberish, a fantasy language that represents a language without any specific meaning of the words. Therefore a Gibberish generator [20] computed output which was randomized, but which had the same statistical distribution of alphabetic characters or combinations of characters given a sample text input. For input we chose a text that represented a standard first time meeting from a textbook for language learning.

Subjects in our study were shown three pairs of videos in alternating order. One pair contained prototypical usage of pauses in speech for typical Asian and Western communication. As the analysis described in Section 3 revealed, the Japanese subjects in our empirical study used significantly more pauses in speech than the German subjects. Accordingly, the simulated dialogs reflecting typical Japanese conversation behaviour should contain more pauses as well. The videos in our preliminary evaluation study lasted half a minute each. Thus, using Table 1 (pauses per

minute) as a basis, the German dialogs in our study contained one pause that lasted for 1 second, whereas the Japanese version contained two pauses that lasted for 1 second and one pause that lasted for 2 seconds. As we did not analyse positioning or function of pauses in speech yet, we placed all pauses between the turns.

Another pair of videos showed culture specific differences in the usage of overlapping speech. Following our analysis described in Section 3 (see Table 2 – overlaps per minute) we integrated one overlap that lasted for 0.3 seconds and two overlaps that lasted for 0.5 seconds into the German dialog. Please note that each agent dialog lasted for approximately half a minute. In contrast the Japanese version contained three overlaps that lasted for 0.3 second, one that lasted for 0.5 seconds and one that lasted for 1 second.

The third pair of videos showed a combination of our findings in the usage of pauses in speech and overlapping speech. Thus, the German video contained one pause that lasted for 1 second, one overlap that lasted for 0.3 seconds and two overlaps that lasted for 0.5 seconds. Accordingly the Japanese version contained two pauses that lasted for 1 second, one pause that lasted for 2 seconds, three overlaps that lasted for 0.3 second, one overlap that lasted for 0.5 seconds and one that lasted for 1 second.

Figure 4. Screenshot of evaluation study, showing a pair of conversation videos displayed by Western-style agents

For a first evaluation, all conversations (German and Japanese), containing different styles of communication management, were displayed with the Western-style characters and rated by German subjects. The agent dialogs were held in German-style Gibberish to avoid that subjects would assume a different cultural background as their own. Apart from communication management, no other aspects of culture-specific communication behaviour were taken into account. To avoid user preferences evoked by differences in nonverbal behaviour the agents in the videos did not exhibit any gestures, culture specific postures. Instead they assumed a neutral posture (standing straight with arms hanging down both sides of their body, see Figure 3). Additionally we showed subjects a neutral video before starting the study. In this video the agents held the same dialog as in the culture-specific videos later but without performing any particular communication behaviour. In this vein, subjects were able to get acquainted with the agents, the scenario and the Gibberish language.

Figure 4 shows a screenshot of one of the three dialog pairs within our evaluation study. We asked subjects to watch the three pairs of videos in alternating order and to judge which one they like better. Therefore subjects could rate their acceptance in three grades on each side, starting from "rather this video than the other one" to "by any means this video".

12 subjects participated in our first evaluation study, 5 female and 7 male. All subjects were German and in an age group between 21 and 28. As we used Western-style characters in our evaluation and invited German subjects to observe and judge the agent dialogs, we expected participants to perceive the German version as more appropriate and than the Japanese version.

Table 3. Overview of user ratings in evaluation study as well as p – values

	Germany	Japan	p
Pause	14	4	< 0.02
Overlaps	16	3	< 0.002
Both	19	3	< 0.002

Table 3 shows an overview of the subjects´ ratings. The values comprise the sum of the 12 subjects' ratings (with preference from 1 to 3 for either the German or Japanese video). The results of our preliminary evaluation study results revealed preferences for the German video versions. Using the two–sided t-test for statistical analysis, we achieved significance for both features, the usage of pauses in speech and overlaps, as well as for their combination (see Table 3 for p–values). Figure 5 graphically shows the mean distribution of the subjects' estimations, where a clear preference for the German version videos can be seen.

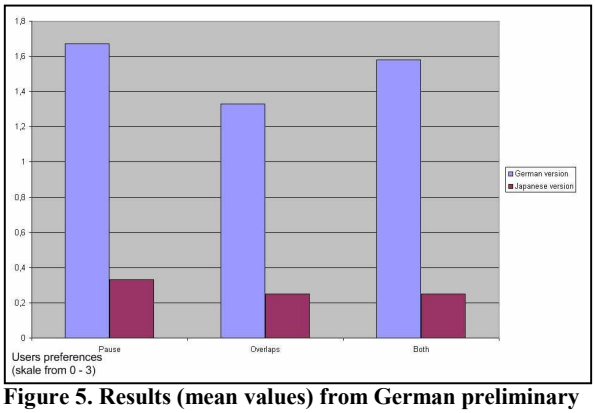

Figure 5. Results (mean values) from German preliminary evaluation study; acceptance on scale from 0 to 3

The results of our preliminary study show that subjects did perceive a difference between culture-specific dialogs that are in line with observations made for their own cultural background and agent dialogs that are not. Additionally subjects were aware why they preferred a certain video and substantiated their perceptions with comments such as: "better timing", "pauses are too long in video x", "agent x cuts agents y off" etc.

5. CONCLUSION AND FUTURE WORK

In this paper, we investigated communication management as one aspect of culture-specific interaction. A corpus analysis for recordings of dialogues with German and Japanese human participants revealed that culture specific dialogue is reflected by a different usage of pauses in speech and overlapping speech. A first empirical study indicates that German subjects seem to prefer communication management in dialogues between virtual agents that is in line with our corpus analysis for human participants. Although our results are very preliminary, we claim that culture related differences in behavior and their integration into multiagent systems is a promising research field, as human users consciously perceive these differences.

In our first evaluation study we tested the impact of different communication management styles on German users. As a next step we also plan evaluation studies with Japanese subjects in collaboration with our Japanese partners. We expect Japanese subjects to prefer the Asian version of communication management. To this point we analyzed the frequency of pauses in speech and overlapping speech and not yet their communicative function. As future work we will conduct a qualitative analysis as well. In particular we want to investigate if e.g. overlapping speech is effectively used for feedback behavior in Japanese conversation and if the corresponding simulation with virtual agents has any impact on human users.

In [19] we described culture related differences in Small Talk behavior focusing on topic selection. In the near future we plan to combine differences in Small Talk behaviour with our findings in communication management behavior. In a long term view, we aim at developing a system that plans dialogs with culture as a parameter and automatically integrates culture specific communication management.

6. ACKNOWLEDGMENTS

The first author of this paper was supported by a Grant from the Elitenetzwerk Bayern (Elite Network Bavaria).
This work was also partly funded by the German Research Foundation (DFG) through its funding for the Cube-G project RE 2619/2-1 and by the European Community (EC) through its funding for the eCIRCUS project IST-4-027656-STP. The authors are solely responsible for the content of this publication. It does not represent the opinion of the European Community which is not responsible for any use that might be made of data appearing therein.

7. REFERENCES

[1] Gudny, R. J., Jonsdottir, Thorisson, K. R. and Nivel, E. 2008. Learning Smooth, Human-Like Turntaking in Realtime Dialogue. In Proceedings of Intelligent Virtual Agents (IVA 08), pages 162-175, Springer Berlin / Heidelberg.

[2] Sato, R., Higashinaka, R., Tamoto, M., Nakano, M., Aikawa, K. 2002. Learning decision trees to determine turn-taking by spoken dialogue systems. In: ICSLP 2002, pp. 861–864.

[3] Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N. 2004. Where to look: a study of human-robot engagement. In IUI '04: Proceedings of the 9th international conference on Intelligent user interface, pages 78–84, New York, NY USA. ACM Press.

[4] Cassell, J. 2000. Embodied conversational interface agents. In Communications of the ACM. Vol. 43, No. 4, April.

[5] Cassell, J., Nakano, Y., Bickmore, T., Sidner, C. L., and Rich, C. 2001. Non-verbal cues for discourse structure. In Proc. Of the 39th Annual Meeting of the Association for Computational Linguistics (ACL).

[6] Nakano, Y. I., Reinstein, G., Stocky, T., and Cassell, J. 2003. Towards a model of face-to-face grounding. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2003), pages 553–561.

[7] Traum, D. and P. Heeman. 1996. Utterance Units and Grounding in Spoken Dialogue. in ICSLP.

[8] Nakanishi, H., Ishida, T., Isbister, K., and Nass, C. 2004. Designing a Social Agent for Virtual Meeting Space. In S. Payr & R. Trappl (Eds.), Agent Culture: Human- Agent Interaction in a Multicultural World (p. 245-266). London: Lawrence Erlbaum Associates.

[9] Cassell, J., Gill, A. & Tepper, P. 2007. Coordination in Conversation and Rapport. Proceedings of the Workshop on Embodied Natural Language, Association for Computational Linguistics . June 24-29, Prague, CZ.

[10] Hofstede, G. 2001. Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage Publications, Thousand Oaks, London.

[11] Hofstede, G.J., Pedersen, P.B. and Hofstede, G. 2002. Exploring Culture - Exercises, Stories and Synthetic Cultures. Intercultural Press.

[12] Ting-Toomey, S. 1999. Communicating Across Cultures, The Guilford Press.

[13] Tromenaars, F., Hampden-Turner, C. 1998. Riding the Waves of Culture, The McGrawn-Hill Companies.

[14] Allwood, J., Nivre, J., & Ahlsén, E. On the Semantics and Pragmatics of Linguistic Feedback, In: Journal of Semantics, 1992, pp. 1-26.

[15] Rehm, M., Bee, N., Endrass, B., Wissner, M. and André, E. 2007. Too close for comfort? Adapting to the user's cultural background. In: 2nd International Workshop on Human-Centered Multimedia.

[16] Kipp, M. 2003. Gesture Generation by Imitation – From Human Behavior to Computer Character Animation. Universität des Saarlandes, PhD. Thesis.

[17] Rochester, S. R. 1973. The Significance of Pauses in Spontaneous Speech. In Journal of Psycholinguistic Research, Vol. 2, No. 1.

[18] Rehm, M., Nakano, Y. and André, E. 2008. Culture-specific first meeting encounters between virtual agents In: Intelligent Virtual Agents 2008, pages 223-236, Springer, Berlin, Heidelberg.

[19] Endrass, B., Rehm, M. and André, E. in press. "Culture - specific Small Talk for Virtual Agents", Proceedings of the Fourth International Workshop on Human-Computer Conversation.

[20] Enevoldsen, K. last viewed September 2008. http://thinkzone.wlonk.com/Gibber/Gibber.htm